

3.3 Substrates for Analytics

Analytics with SNOMED CT may be deployed on a wide range of data sources as summarized in the table below. These data sources are also known as the 'substrate' of the analytics. Please note that data which is not natively coded using SNOMED CT may be transformed using one of the techniques described in section 5 [Preparing Data for Analytics](#). These techniques may be used to transform heterogeneous data recorded using free text or a variety of code systems into SNOMED CT, which can serve as a common reference terminology for analysis.

Table 3.3-1: Direct and indirect substrates for SNOMED CT based analytics

Analytics Substrate	Examples	Coding	Information Model
Unstructured free text document	Dictated clinical letter	Natural language	None or informal headings
	Typed discharge summary letter		
Structured documents with free text fields	Assessment form	Natural language	Standardized headings and fields
	Discharge summary form		
Structured documents with free text and post-coded classification (i.e. added by clinical coders after the clinical event)	Discharge summary form with post-coded classification	Classifications (e.g. ICD)	Formal information model (typically simple)
Structured documents with non-SNOMED CT coding (e.g. proprietary, local or other coding system)	Standalone clinical application using departmental codes	Local code system, controlled vocabulary or legacy clinical terminology	Formal information model
	Enterprise-wide healthcare system using local dictionaries and pick-lists		
	Electronic patient record using regional coding system (such as UK Primary Care systems)		
Structured documents with SNOMED CT content	Cardiology report	SNOMED CT	Formal information model
	GP event summary		
'Big data' data store	Data warehouse	Various coding systems	Mixture of both structured and unstructured data
	Data store containing a mixture of substrates		